

UNIVERSITÁ DEGLI STUDI DI MILANO-BICOCCA

**Corso di laurea Magistrale in Data Science
Social Media Analytics**



Twitter Social Analysis nel dibattito vaccino-GreenPass

Relazione ad opera di:

Davide Mobilia, 805785

Eric Spinelli, 799827

Anno 2021/2022

Abstract

Questo progetto si pone l'obiettivo di analizzare il network creatosi sui social, in particolare Twitter, dalle discussioni intorno all'argomento covid. Nello specifico abbiamo analizzato i tweet riguardanti il tema vaccino e green pass e come le diverse communities interagiscano tra loro.

L'analisi sarà quindi incentrata sulle posizioni contrapposte tra chi è favorevole ai vaccini e chi, come i no vax, li vede come un'imposizione e critica le relative misure prese dal governo.

Come ulteriore analisi si sono inoltre estratte le interazioni tra questi due principali cluster e alcune correnti politiche. L'obiettivo di questa seconda analisi è verificare se alcune posizioni di natura politica e sociale siano più affini a espressioni novax piuttosto che pro-vaccino e greenpass.

Dati e raccolta

La raccolta è stata fatta su due hashtag: #NoGreenPass" e "#vaccinatevi". Questo ha permesso di identificare con chiarezza le diverse posizioni rispetto al tema affrontato, da un lato persone contrarie a vaccini e restrizioni dall'altro coloro che sono favorevoli e spingono a vaccinarsi. Una delle principali difficoltà per l'identificazione di due cluster distinti è stata proprio la ricerca di hashtag e parole chiave che potessero identificare chiaramente una specifica community.

Ricerche come #novax, #covid19 etc portavano a risultati ambigui e non facilmente riconoscibili. La maggior parte degli hashtag sono infatti comuni denominatori per la maggior parte dei tweet, di una o dell'altra categoria.

A complemento di questa raccolta abbiamo selezionato alcuni hashtag e menzioni che potessero rappresentare forze politiche o movimenti di spicco legati in qualche modo a campagne pro/contro vaccino, in particolare: "@CasaPoundItalia", "@matteosalvinimi", "@EnricoLetta".

La raccolta è stata effettuata il giorno 30 dicembre 2021, in corrispondenza all'emanazione del nuovo decreto sul Super Green Pass, allo scopo di poter analizzare le reazioni da parte delle diverse posizioni.

Per ciascun hashtag o menzione si sono estratti 5000 tweet sfruttando l'API di Twitter. Per la creazione del grafo si è scelto di studiare le interazioni tra gli utenti sulla base dei retweet effettuati. Per ottenere questo risultato si è incentrata l'analisi sugli utenti, e quindi i tweet e i ritweet raccolti sono stati aggregati per utente.

In totale si sono individuati 5634 nodi, corrispondenti a tutti gli utenti coinvolti in interazioni di retweet quindi, coloro che hanno effettuato un tweet e coloro che lo hanno ritwittato. Tra questi utenti si sono individuati 9273 archi, ciascuno dei quali in corrispondenza di un retweet effettuato. Da notare quindi che nella rete estratta, alcuni utenti che sono stati ritwittati, hanno a loro volta ritwittato altri utenti presenti nella rete. Al contempo, molti utenti hanno ritwittato più utenti presenti nella stessa rete.

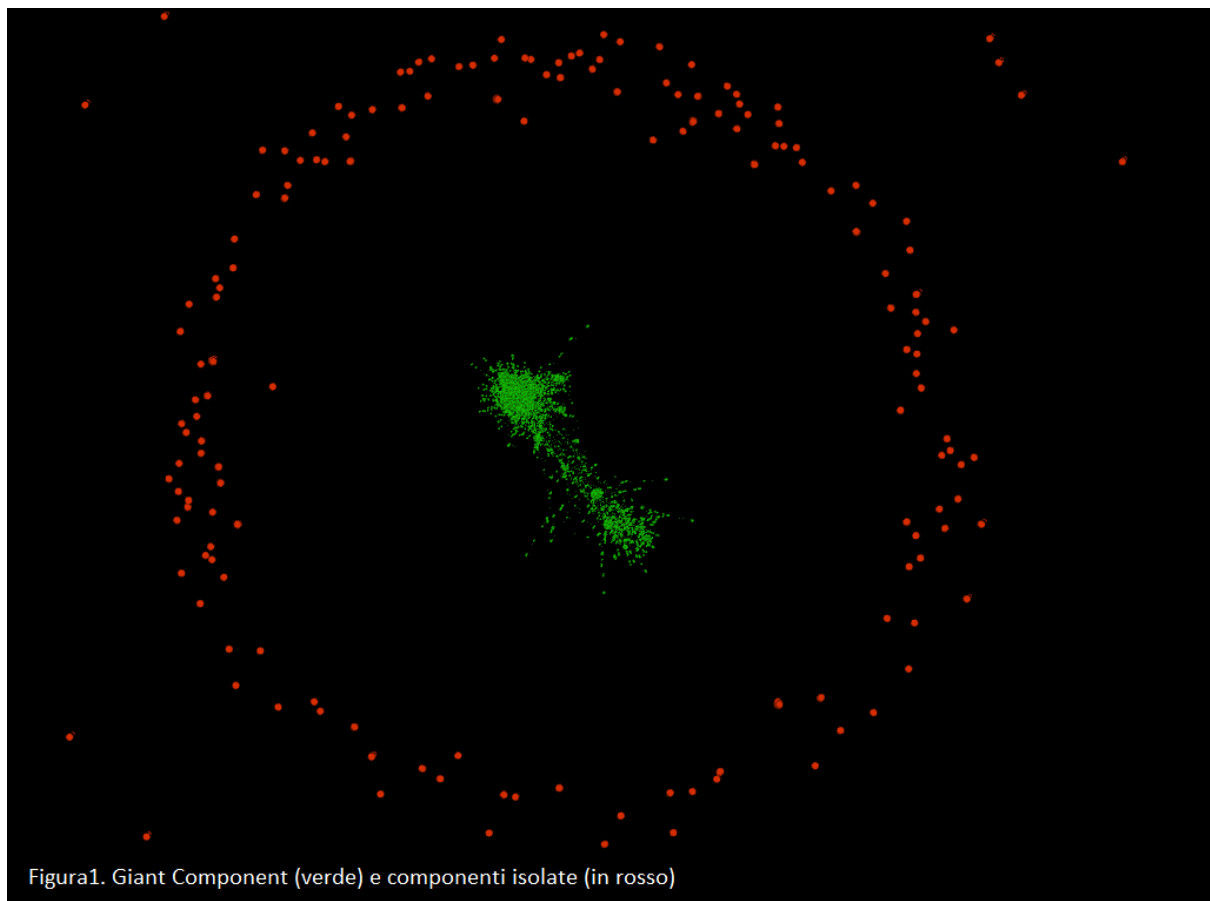
Analisi grafo

I due cluster principali su cui ci concentreremo per la nostra ricerca sono le persone contrarie a vaccino e green pass e coloro che vedono nelle misure restrittive e nel vaccino la soluzione da percorrere. Pur avendo raccolto solo l'hashtag #NoGreenPass per la prima categoria e #vaccinatevi per la seconda, assumeremo che le due raccolte abbiano il significato più ampio prima descritto. Per comodità, in seguito, verrà utilizzato il termine novax per le persone contrarie a vaccini e misure restrittive identificate nella nostra raccolta da #NoGreenPass mentre indicheremo con utenti pro-vaccino la categoria opposta e identificata nella raccolta con l'hashtag #vaccinatevi.

Il grafo creato dai retweet delle diverse raccolte è composto da 5634 nodi e 9273 archi, in particolari i nodi sono così raggruppati:

#NoGreenPass	1635
#vaccinatevi	1304
@CasaPoundItalia	299
@EnricoLetta	1147
@matteosalvinimi	1249

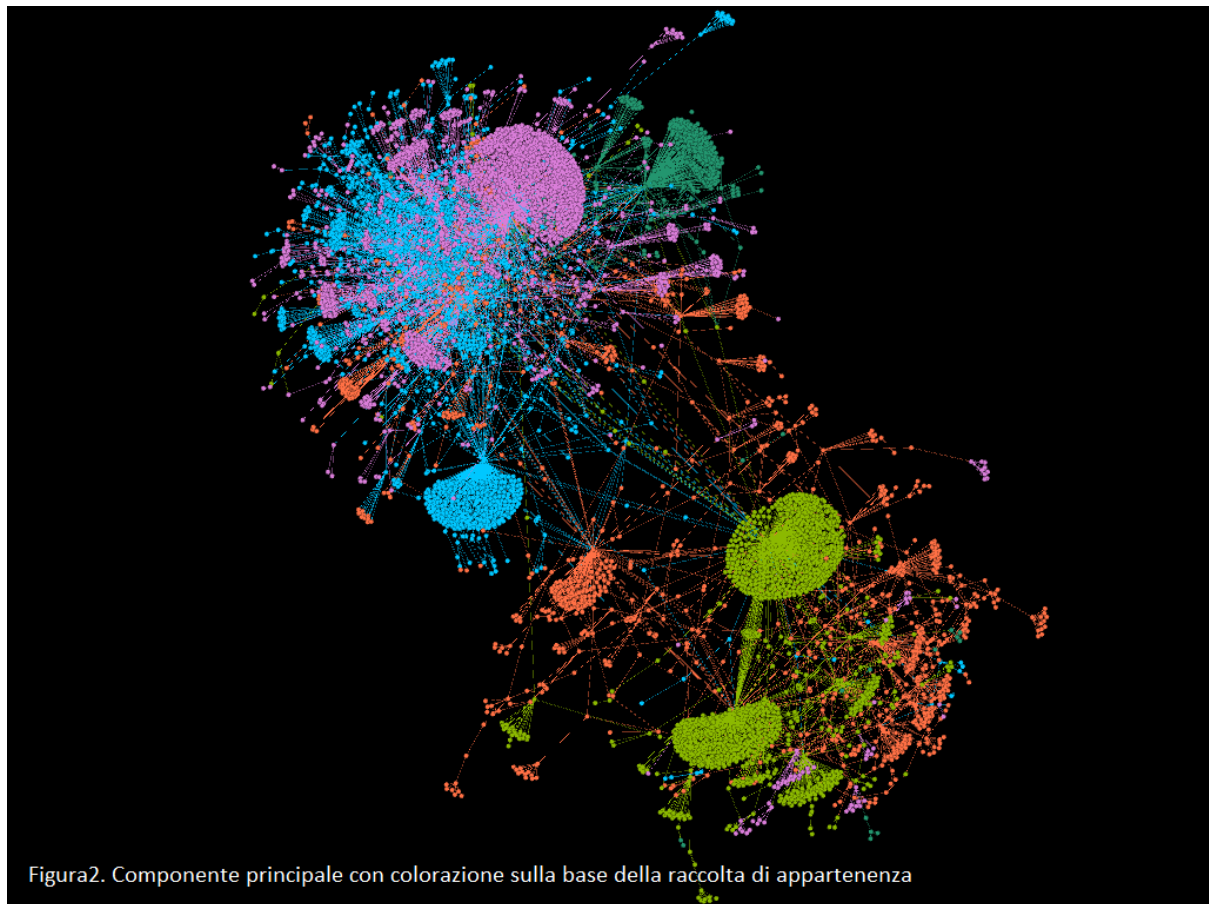
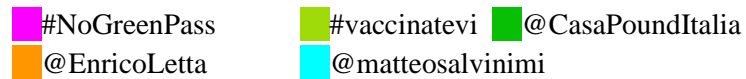
Durante l'analisi della rete si sono però subito incontrate difficoltà a causa di numerosi nodi collegati solo con altre poche unità. Queste componenti indicano quindi la presenza di utenti che sono stati ritwittati solo da un numero esiguo di persone e rimasti quindi scollegati dalla rete (vedi Figura1).



Si è deciso quindi di isolare la giant component e procedere con le analisi su questa nuova rete composta da 5190 nodi, le proporzioni dei diversi gruppi rimangono le stesse.

Andando a concentrarci sul giant component, cerchiamo di capire le caratteristiche della nostra rete partendo da una diversa visualizzazione che possa esaltarne i differenti cluster.

In Figura2 la rappresentazione della Giant Component con colorazione sulla base delle diverse raccolte:



Dal grafo si possono identificare due principali raggruppamenti che raccolgono #NoGreenPass, @CasaPoundItalia e @matteosalvinimi da un lato e #vaccinatevi, @EnricoLetta dall'altro.

Le diverse raccolte sembrano tutte orientarsi intorno ad un proprio nodo centrale, in verde ne notiamo ben due.

Un comportamento differente caratterizza le componenti associate a @EnricoLetta, per cui non si riescono a identificare nella rete raggruppamenti significativi.

Ulteriori analisi tra i diversi cluster, nodi e loro eventuali interazioni saranno approfondite successivamente attraverso la degree analysis e la community detection.

Per avere un'idea generale della struttura della nostra rete abbiamo raccolto in una tabella alcuni dei principali indici per la valutazione dei grafi:

Assortativity	Edge Connectivity	Node Connectivity	Density	Radius	Diameter	Overall reciprocity
-0.2152	1	1	0.0005	7	14	0.0035

L'Assortativity è bassa e di valore negativo, questo ci fa propendere per un caso di rete disassortativa. Ci aspettiamo quindi che gli HUB tendano a connettersi a piccoli nodi piuttosto che ad altri HUB e che la loro rimozione creerebbe un grosso danno alla rete stessa. Caratteristiche in linea con una rete costruita su retweet dove gli utenti più popolari ne determinano le caratteristiche della stessa, struttura evidente anche dalla rappresentazione grafica precedentemente mostrata.

La density molto bassa indica che siamo molto distanti da un grafo completo, ovvero una rete in cui tutti i nodi sono connessi tra loro. La maggior parte dei nodi avrà infatti poche connessioni e perlopiù entranti negli HUB rappresentanti gli utenti più ritwittati. Infine l'overall reciprocity ci indica che quasi nessun utente ha contraccambiato il retweet ricevuto.

Degree Analysis (HUBS & Spam Users)

Studiamo ora la distribuzione dei gradi all'interno della nostra rete e le caratteristiche di quei nodi che si distinguono per l'alta numerosità di archi entranti (HUBS) o uscenti (utenti più attivi).

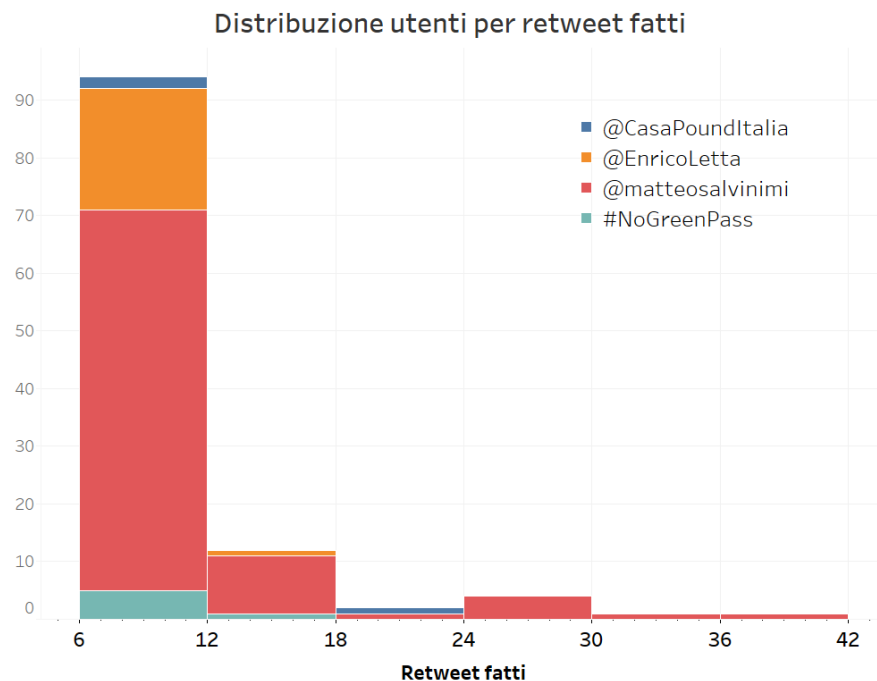
	Min	Max	Avg	Grado=0
Out-degree	0	38	1.40	9.2%
In-degree	0	1080	1.40	86.1%

Queste prime misure confermano i risultati trovati in precedenza: la nostra rete è caratterizzata da nodi mediamente poco connessi. La maggior parte di essi sono utenti che non vengono ritwittati ma che al contrario ritwittano utenti influenti.

Spam Users

Andiamo ora ad analizzare nello specifico la distribuzione di quegli utenti che risultano più attivi, che hanno quindi effettuato il maggior numero di retweet.

Per individuare tali utenti andremo ad escludere i nodi che hanno un numero di archi uscenti (retweet fatti) ≤ 5 , in questo modo escludiamo il 97.8% degli utenti non influenti alla ricerca. La distribuzione del rimanente 2.2%, che avrà quindi effettuato almeno 6 retweet, è stata plottata attraverso un istogramma e suddivisa sulla base dell'hashtag o menzione di riferimento.



Viene subito reso evidente come la maggior parte di questi utenti appartenga alla raccolta @matteosalvinimi mentre non compare nella statistica l'hashtag #vaccinatevi.

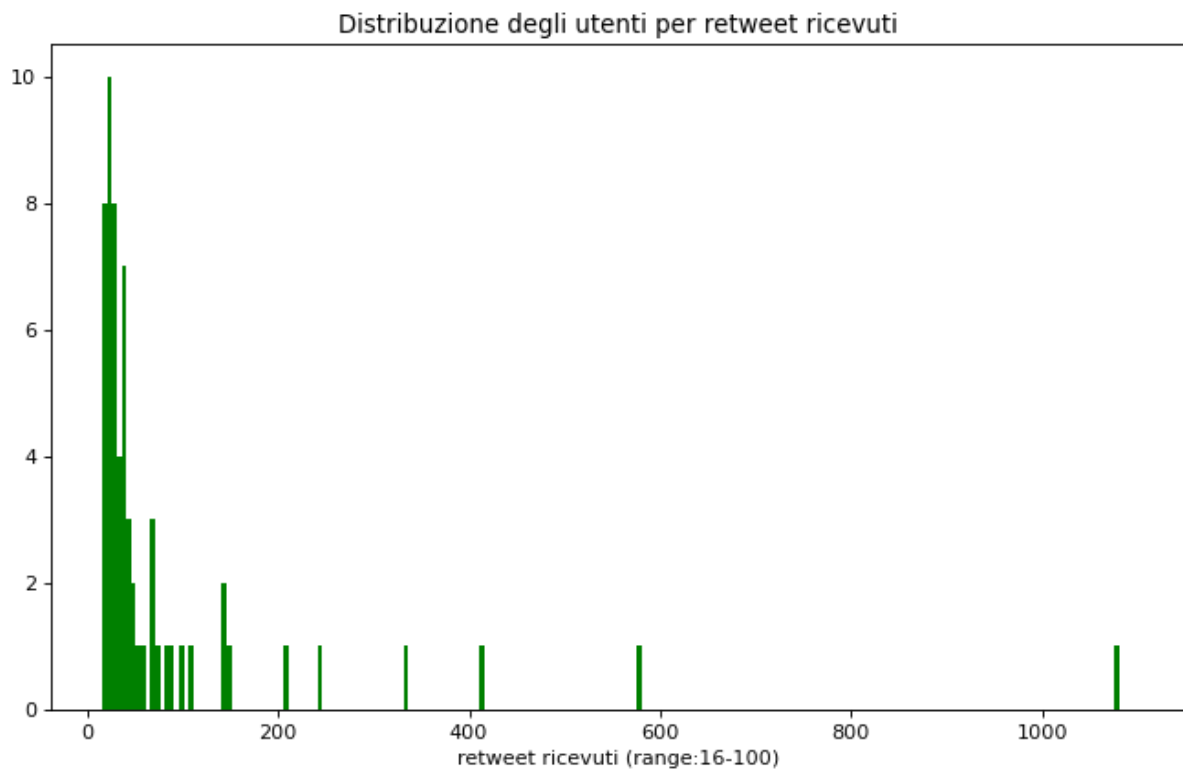
Andando ad approfondire ulteriormente questo aspetto si è calcolato la percentuale di appartenenza dei top30 utenti più attivi (range dei retweet:10-38). I dati confermano i risultati della distribuzione appena osservata: l'80% dei top30 utenti più attivi sono legati a @matteosalvinimi. A seguire abbiamo #NoGreenPass con il 10%, @EnricoLetta al 6.66% e @CasaPoundItalia 3.33%

Pur avendo delle classi bilanciate (escludendo la raccolta @CasaPoundItalia), risulta evidente una grande sproporzione dell'attività social per gli utenti associati a @matteosalvinimi rispetto agli altri gruppi. La spiegazione più immediata di questo fenomeno potrebbe essere dovuta ai grandi investimenti e la propaganda che Salvini e il suo partito operano costantemente sui social. Ulteriore dato di interesse è quello per cui l'hashtag #vaccinatevi non è presente in queste distribuzioni, gli utenti pro vaccino sembrerebbero quindi essere meno attivi rispetto alla loro controparte.

HUBS

Andiamo ora ad analizzare la distribuzione e le caratteristiche dei nodi con il maggior numero di archi entranti, ovvero quegli utenti che hanno ricevuto il maggior numero di retweet.

A questo scopo andiamo a selezionare solo l'1.2% dei nodi, ovvero quegli utenti che abbiano ricevuto almeno 15 retweet per analizzarne la distribuzione.



Attraverso questa rappresentazione possiamo identificare in linea generale le dimensioni degli HUB (in termini di retweet) e quanti siano quelli più significativi. Con un numero di retweet superiore a 200 si identificano 6 diversi utenti che si differenziano maggiormente rispetto ai rimanenti nodi.

Utente	Retweet ricevuti	In Degree Centrality	Hashtag
<i>Lorenzo62752880</i>	1080	0.208	#NoGreenPass
<i>mariannaaprile</i>	579	0.112	#vaccinatevi
<i>Cartabellotta</i>	415	0.080	#vaccinatevi
<i>matteosalvinimi</i>	332	0.064	@matteosalvinimi
<i>CasaPoundItalia</i>	245	0.047	@CasaPoundItalia
<i>gianluca826</i>	207	0.040	@matteosalvinimi

Come si può osservare gli utenti più significativi provengono da diverse raccolte, mancando solamente @EnricoLetta nella classifica.

Al primo posto spicca per numero di retweet un utente seguito sui social per i suoi post novax. In particolare il commento oggetto della nostra raccolta critica duramente la categoria dei giornalisti in quanto complici di un attacco alle libertà personali dei cittadini.

Interessante la contrapposizione con il secondo utente con più retweet: Marianna Aprile, giornalista pro-vaccino il cui post cita un articolo di giornale riguardante i dati sull'efficacia dei vaccini stessi. A seguire l'immagine che contrappone i due diversi tweet di natura totalmente opposta.



Il filosofo Giorgio Agamben:

"Gli storici guarderanno a questo momento storico evidenziando come i giornalisti abbiano dato prova della più vergognosa e infame complicità".

[#NoGreenPass](#)

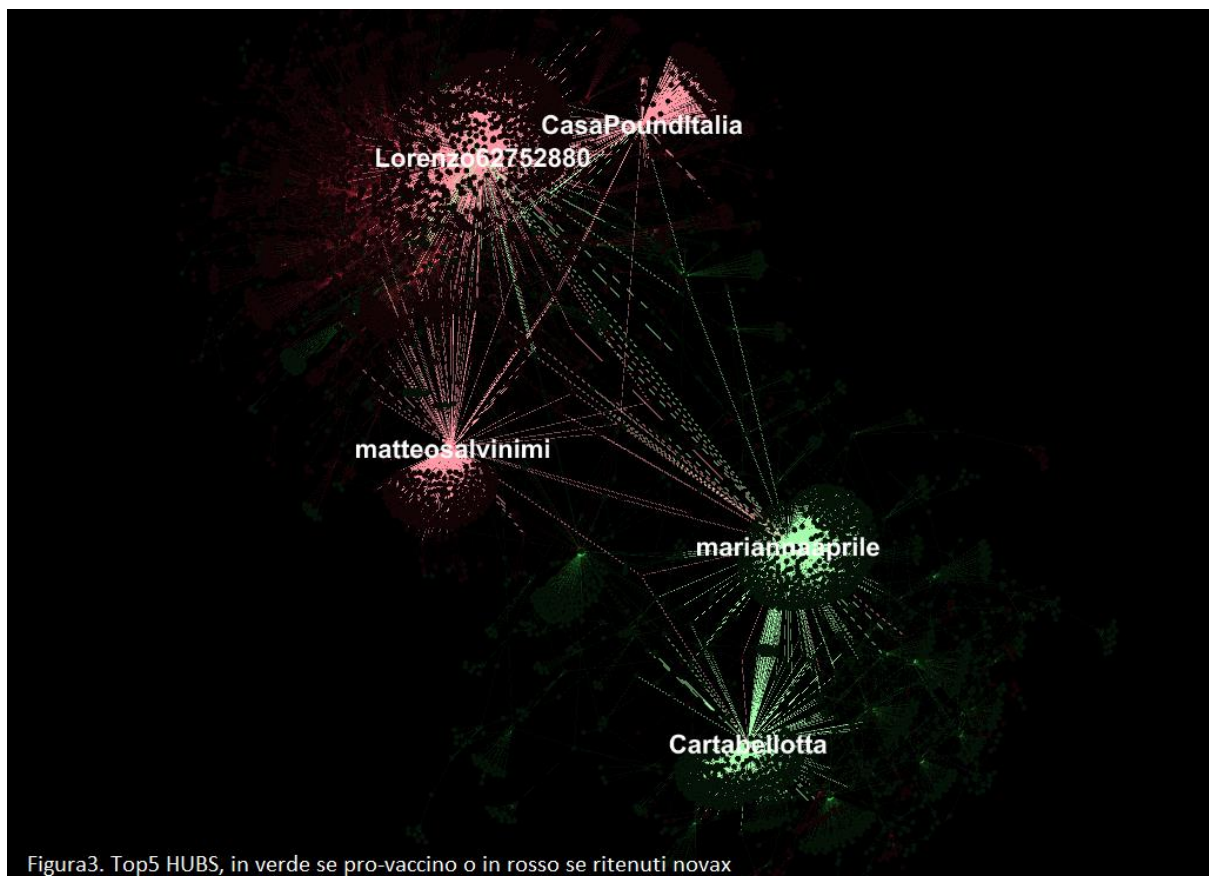


Se proprio dovete dare i numeri sui vaccini, provate con questi

[#vaccinatevi](#)

I picchi di contagio a confronto			
23 dicembre 2021		13 novembre 2020	
44.595	8.722	40.902	30.914
Contagi I casi registrati ieri a fronte di 901.450 tamponi	Ricoveri ordinari Nei reparti Covid ordinari ci sono 178 malati in più	Contagi I nuovi casi registrati un anno fa quando non c'erano i vaccini	Ricoveri ordinari I pazienti ricoverati con sintomi erano in costante crescita

Per cogliere meglio le loro interazioni mostriamo ora una rappresentazione grafica dei primi 5 HUB, colorati sulla base del gruppo di appartenenza: novax o pro-vaccino.



Considerando i valori della *in_degree centrality* *Lorenzo62752880* raccoglie ben il 20.8% di tutti i retweet della rete mentre *mariannaaprile* l'11.2%. Considerando i top5 HUB della rete mostrati in Figura3. raggiungiamo il 51.1% dei retweet totali, questo a dimostrazione di come la nostra rete sia di fatto costruita intorno a pochi utenti influenti.

Ulteriore conseguenza di questo aspetto sono i risultati della *closeness centrality*: *Lorenzo62752880* e *mariannaaprile* risultano avere i più alti valori di questo indice, rispettivamente 0.2 e 0.14. La *closeness centrality* essendo calcolata sulla base dello *shortest path* sarà fortemente condizionata da quei nodi che hanno ricevuto direttamente il maggior numero di retweet.

La *betweenness centrality* invece ha mediamente valori molto bassi, il più alto valore trovato è pari a 0.000103. Trattandosi di un grafo diretto ci aspettiamo infatti una scarsa circolazione delle informazioni e l'analisi della *betweenness centrality* non ha quindi una particolare rilevanza all'interno della nostra rete.

Community Detection

A seguito delle analisi sul grafo e delle relative osservazioni, per procedere con la community detection siamo partiti da un assunto iniziale, sostenuto dall'osservazione dei comportamenti dei partiti presi in considerazione relativamente al tema no vax, interni ed esterni a Twitter. La nostra ipotesi prevede che nel nostro network si individuano 2 principali community: una pro vax che oltre all'hashtag #vaccinatevi include la menzione @EnricoLetta, rappresentante PD di centrosinistra; una no vax che oltre all'hashtag #NoGreenPass include @matteosalvinimi, rappresentante Lega di centrodestra e @CasaPoundItalia, movimento politico di estrema destra.

In questo paragrafo ci poniamo l'obiettivo di verificare a livello grafico e di clusterizzazione quanto questo assunto rispecchi la realtà e in che modo le diverse communities interagiscono tra loro. In generale, come ci si aspetta, gli utenti hanno una tendenza a ritwittare e ad essere ritwittati da utenti della stessa categoria, rendendo due i raggruppamenti ben distinti.

Nella seguente visualizzazione si sono raggruppati gli utenti relativi ai propri hashtag secondo la logica appena descritta. Si può notare come la componente novax sia molto più numerosa e compatta rispetto agli utenti pro-vaccino. Si può comunque notare come siano presenti utenti che pur appartenendo ad una categoria interagiscono, attraverso retweet, con quella contrapposta.

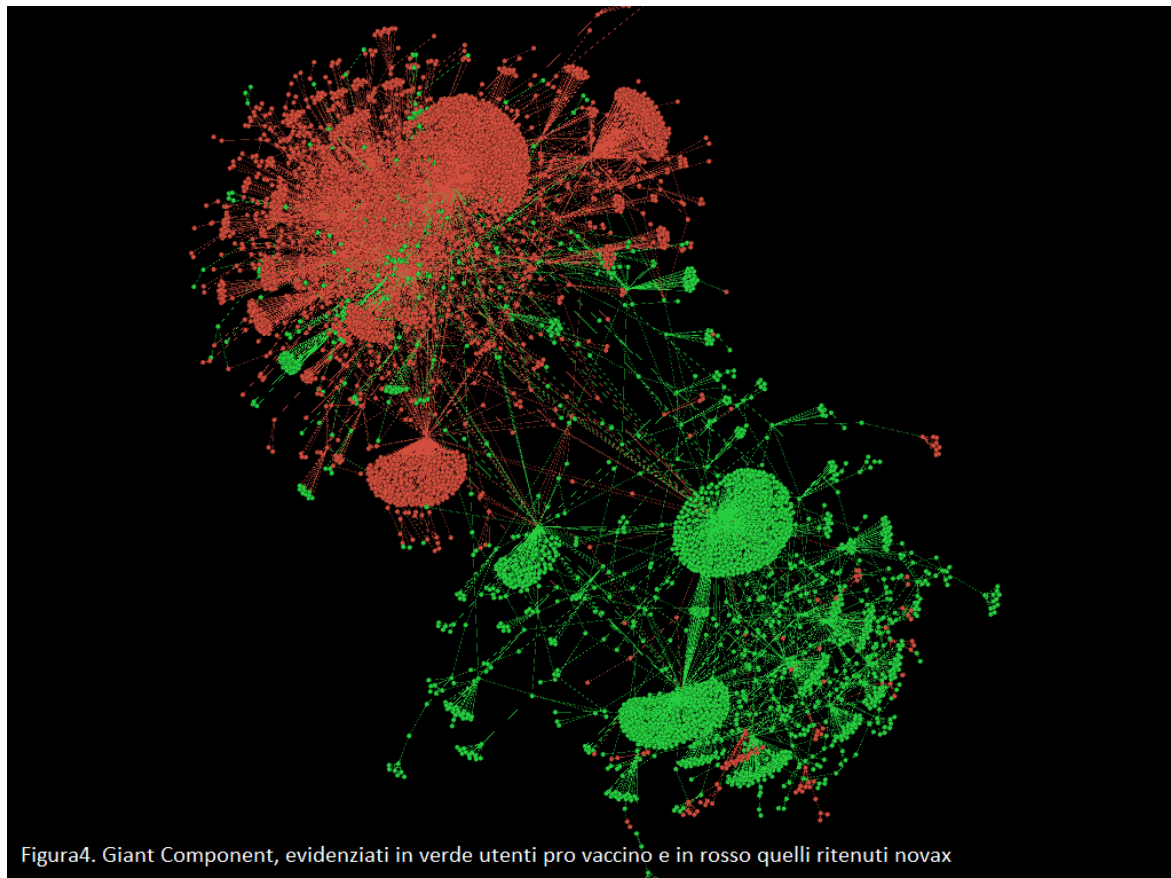


Figura4. Giant Component, evidenziati in verde utenti pro vaccino e in rosso quelli ritenuti novax

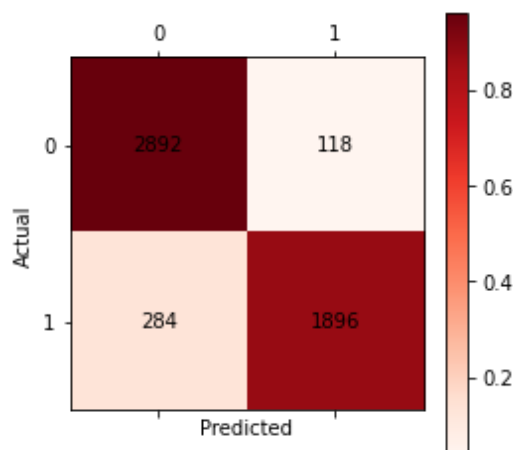
Allo scopo di testare la nostra ipotesi iniziale si sono confrontati alcuni algoritmi di community detection. I modelli messi a confronto sono stati: Newman's leading eigen vector, Kmeans e Girvan-Newman.

Di seguito si descriverà l'algoritmo che ha prodotto il risultato migliore, cioè Girvan-Newman. Si tratta di un algoritmo iterativo che ad ogni iterazione elimina gli archi che presentano il numero più alto di cammini brevi tra i nodi che connessi attraverso di essi, ovvero quelli più importanti per le connessioni tra i nodi interne al network. L'algoritmo si basa quindi sulla betweenness centrality. L'idea di fondo è appunto quella di individuare quegli archi interni al network che hanno betweenness centrality più alta. Una volta eliminati questi archi, risulteranno più evidenti i raggruppamenti dei nodi che individuano le rispettive community.

Come presentato dallo pseudo codice, le iterazioni proseguono per tutti gli archi del grafo. Ne consegue che il costo computazionale dell'algoritmo è particolarmente elevato e ciò è stato dimostrato dal fatto che nel nostro caso l'esecuzione dell'algoritmo ha impiegato oltre 7 ore a giungere al termine.

```
REPEAT
  LET n BE number of edges in the graph
  FOR i=0 to n-1
    LET B[i] BE betweenness centrality of edge i
    IF B[i] > max_B THEN
      max_B = B[i]
      max_B_edge = i
    ENDIF
  ENDFOR
  REMOVE edge i FROM graph
UNTIL number of edges in graph is 0
```

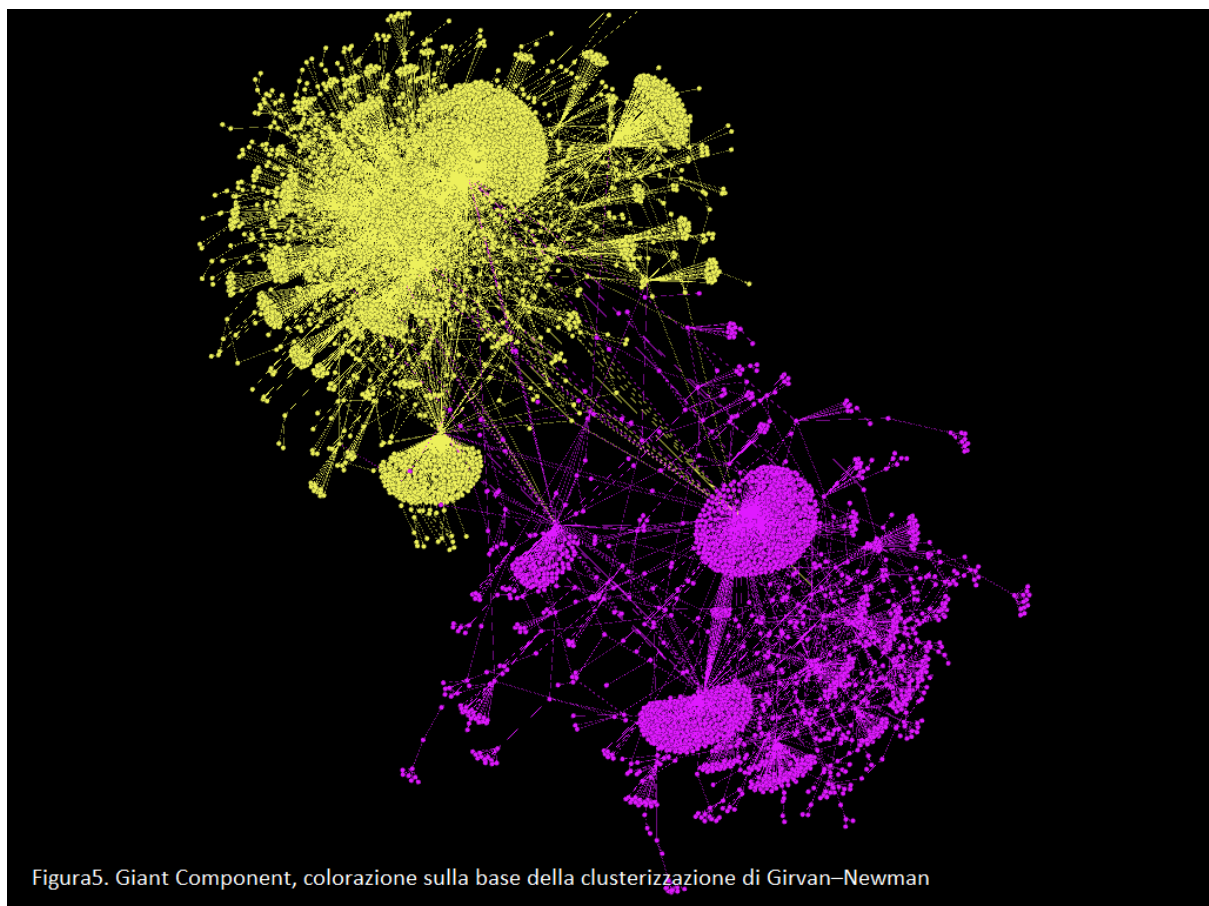
Nonostante il drawback dell'elevato costo computazionale, la valutazione delle performance dell'algoritmo ci ha spinto a selezionarlo come migliore, raggiungendo un'Accuracy del 92.3%. Di seguito si presenta la confusion matrix che sintetizza i risultati della classificazione, considerando come 0 la classe novax, mentre come 1 la classe pro-vax:



Rispetto alle misure di performance quali Precision, Recall, F1-Score, il modello dimostra la sua robustezza:

	Precision	Recall	F1-Score
<i>No Vax</i>	0.911	0.961	0.935
<i>Pro vax</i>	0.941	0.87	0.904

In ultimo si presenta il grafo risultante per l'individuazione delle due community con l'applicazione di Girvan-Newman. A seguito dell'analisi sulle performance del modello e della visualizzazione del risultato possiamo dire che la nostra ipotesi di partenza è validata. Si può quindi inferire che il Partito Democratico, di centro sinistra, sia ideologicamente più affine alle idee pro vax (community viola); di contro, il partito Lega, di centro destra e il movimento di estrema destra Casapound, sono sostenitori delle idee no vax (community gialla).



Sentiment Analysis

Sulla raccolta di tweet effettuata, si è eseguita la sentiment analysis, allo scopo di individuare il sentimento diffuso tra le idee delle due community individuate.

Per raggiungere questo obiettivo, per prima cosa si sono preprocessati i testi dei tweet raccolti. Questo è stato fatto allo scopo di rendere i testi più puliti possibile prima di essere processati dal modello per il sentiment score e di preparare i token per i wordcloud.

Di seguito si dettagliano le fasi di preprocessing:

1. Lowercase: si sono convertite tutte le maiuscole in minuscole
2. Pulizia da simboli: si è effettuata la tokenization e si sono rimossi tutti gli hashtag, menzioni e indirizzi web presenti nel testo. Inoltre si sono rimossi tutti i simboli che non fossero testo applicando la punctuation removal. A questo livello del preprocessing si è creata una nuova colonna nel nostro dataframe di lavoro per mantenere le emoticon nel testo, sfruttate in seguito dal nostro modello per performare la sentiment analysis, grazie a un lexicon destinato all'individuazione delle emoji emoji_utf8_lexicon. La rimozione completa dei simboli è invece servita per il wordcloud.
3. Stopwords removal: si è estratta una collezione di stopwords sfruttando il pacchetto nlt e si sono rimosse tutte le stopwords dal testo
4. Lemmatization: come pacchetto per performare la lemmatizzazione si è sfruttato Spacy, per il quale si è scaricato e applicato un dizionario di italiano esteso, che il modello ha sfruttato per la trasformazione dei termini nel loro lemma

Per performare la sentiment analysis si è scelto Vader, un modello lexicon based. In particolare, nella scelta della sua applicazione dopo una valutazione empirica delle performance rispetto a nltk, si è scelto l'utilizzo del pacchetto vaderSentiment. Dal momento che i nostri testi sono in lingua italiana, abbiamo integrato il lexicon di base con il lexicon Sentix, uno dei più completi per la lingua italiana, normalizzando gli score rispetto alla metrica del dizionario già presente. Infatti gli score di vader_lexicon hanno un range compreso tra -4 e +4, mentre quelli di sentix variano tra -1 e +1.

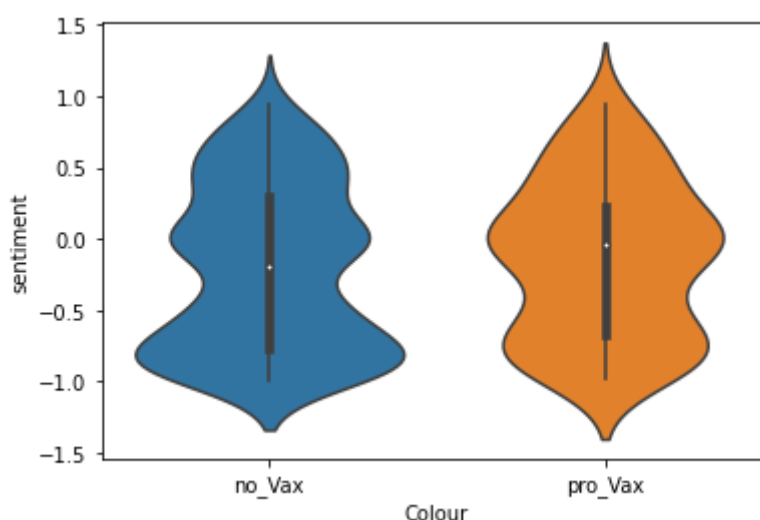
Nella nostra applicazione ci siamo trovati di fronte ad un caso particolare, ovvero quello dell'attribuire la giusta valenza ai termini *positivo* e *negativo*, dal momento che come ben si sa, riguardo al tema del Covid possono assumere un significato diverso da quello di uso comune. A seguito di un'analisi dei tweet a nostra disposizione, abbiamo scelto di attribuire al termine *positivo* uno score nettamente negativo, visto che la maggior parte dei tweet con questo termine assumevano una linea d'accusa o di ammonimento. Al termine *negativo* invece abbiamo attribuito un valore neutro, dal momento che è reso effettivamente ambiguo dipendentemente dal contesto.

Il risultato dell'algoritmo di sentiment Vader che abbiamo preso in considerazione è il compound score, calcolato come somma degli score positivi, negativi e neutri normalizzata tra -1 e +1. Di seguito si riportano delle statistiche di base legate al sentiment score sui testi dei tweet:

	Min	Max	Avg	Sd
<i>Pro Vax</i>	-0.979	0.949	-0.146	0.542
<i>No Vax</i>	-0.990	0.937	-0.210	0.574

Come si può notare la differenza tra le statistiche della sentiment rispetto ai due gruppi non è netta. Tuttavia, si può notare che c'è una tendenza più negativa nel gruppo no vax. Questo si evince sia dalla media, sia dal minimo il cui score sfiora l'estremo negativo. Le emozioni mediamente negative sono possono essere imputabili all'argomento molto sentito e delicato quale quello del covid. Ulteriore spiegazione potrebbe essere legata al fatto che un gran numero di tweet esprime indignazione nei confronti delle idee e delle posizioni della community opposta.

Dal violin plot presentato di seguito si conferma quanto appena detto, denotando una negatività pervasiva in entrambe le community. Tuttavia si può notare come la community no-vax abbia una maggiore densità verso nei valori più bassi di sentiment.



Di seguito si mettono a confronto il testo del tweet più positivo:

“@roberto2_lamela #Poverementi #novax,il vaccino protegge non facendo miracolosam. sparire il virus,ma rinforzando le difese immunitarie in modo mirato,ti puoi infettare,ma hai già dentro di te buone armi per combatterlo e non ammalarti.+vaccinati= -terreno fertile per il virus.#vaccinatevi 🙏”

con quello del tweet più negativo

“Ma come c@zzo è possibile soltanto pensare che una persona possa morire dopo aver fatto un' iniezione e dire che hai un sospetto. È una certezza, chiaro?! #NoGreenPass #NoDittatura #giulemanidaibambini #governodeipeggiori”

Come si può notare il tweet più positivo è relativo all’hashtag #vaccinatevi, mentre quello più negativo contiene l’hashtag #NoGreenPass. Si può inoltre notare come il tweet più positivo, tuttavia, non esprima una sentiment nettamente positiva, a differenza di quello negativo che esprime esplicitamente un sentimento di rabbia. Questo a dimostrazione del fatto che c’è una tendenza da parte della maggior parte dei tweet della raccolta ad esprimere negatività.

In ultimo si presentano i wordcloud. Il primo, relativo alla community pro vax, mette in risalto il principale motivo di preoccupazione legato al covid, ovvero il “contagio”. Altre parole relative al tema sono “caso” rispetto ai casi di covid, “morire” relativamente ai decessi. Altri termini come “servire” e “tampone” sembrano avere una numerosità simile, facendone trasparire una connessione.



Passando alla community no vax, prevalgono i termini “ancora”, “altro”, “no”, lasciando intendere una nota forte di dissenso, e del ripetersi sempre delle stesse circostanze “senza” che si trovi un’altra soluzione. Un’altra parola che risalta è “libertà”, quella in cui la community no vax si caratterizza dal sentirsene privata.



Valutazioni e sviluppi futuri

Lo sviluppo di questo progetto ci ha permesso di identificare le diverse community e posizioni che ruotano all'interno del dibattito pubblico sulla delicata questione covid e delle sue ripercussioni politiche, sociali ed economiche.

Come nella maggior parte dei social le community e i dibattiti si creano intorno a poche personalità che ne diventano quindi influencer.

In prima istanza emerge la contrapposizione dicotomica tra novax e utenti pro-vaccino, con caratteristiche ben distinte. All'interno del primo gruppo le personalità che emergono sono generalmente cittadini anonimi che trovano all'interno dei social lo sfogo necessario e causato da situazioni di disagio e malessere. In contrapposizione, tra gli influencer pro-vaccino troviamo personaggi che cercano di sfruttare il loro ruolo e la professione per esprimere le proprie posizioni e convinzioni. Tra questi abbiamo trovato infatti una giornalista (Marianna Aprile) e un medico (Cartabellotta) i cui post cercavano di diffondere informazioni e dati che nella loro visione misuravano al meglio la realtà.

In seconda analisi si è visto come anche l'affiliazione politica influisca sulle posizioni prese e condivise all'interno dei social. Personalità e movimenti di destra quali Salvini e CasaPound hanno una visione molto vicina se non sovrapponibile al mondo novax. Parallelamente personaggi più moderati e di sinistra quale Enrico Letta hanno più interazioni e posizioni favorevoli alle idee pro-vaccino.

Ben consapevoli che queste analisi siano solo uno spunto di riflessione da cui partire, i risultati ottenuti costituiscono una buona introduzione ad eventuali approfondimenti.

Quali altri interessi e posizioni accomunano gli utenti dei due diversi cluster?

C'è una reale e netta distinzione in quanto a presenza social tra le due tipologie utente? Abbiamo ad esempio osservato un'intensa attività degli utenti che si riferivano a Salvini, un'analisi di bot detection potrebbe creare ulteriori spunti di riflessione su questo tema.

In definitiva oltre alle numerose informazioni che questo studio ci ha permesso di raccogliere, quello che questo progetto ci ha permesso di capire è la vastità e complessità dell'argomento trattato e sulle numerose diramazioni che questa ricerca potrebbe prendere.

Riferimenti

- Slide del corso di Social Media Analytics, a cura di Marco Viviani e Elisabetta Fersini
- https://memgraph.com/blog/community_detection-algorithms_with_python_networkx
- <http://valeriobasile.github.io/twita/downloads.html>
- <https://developer.twitter.com/en/docs/twitter-api>
- <https://medium.com/analytics-vidhya/girvan-newman-the-clustering-technique-in-network-analysis-27fe6d665c92>
- <https://www.analyticsvidhya.com/blog/2020/04/community-detection-graphs-networks/>
- VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text C.J. Hutto
Eric Gilbert